COLORADO SCHOOL OF MINES
EARTH • ENERGY • ENVIRONMENT

NREL
NATIONAL RENEWABLE ENERGY LABORATORY

To Buy or Lease Solar PV: A Selection Bias Problem

Jacquelyn Pless, *Research Economist, Joint Institute for Strategic Energy Analysis at NREL, and PhD student, Colorado School of Mines*

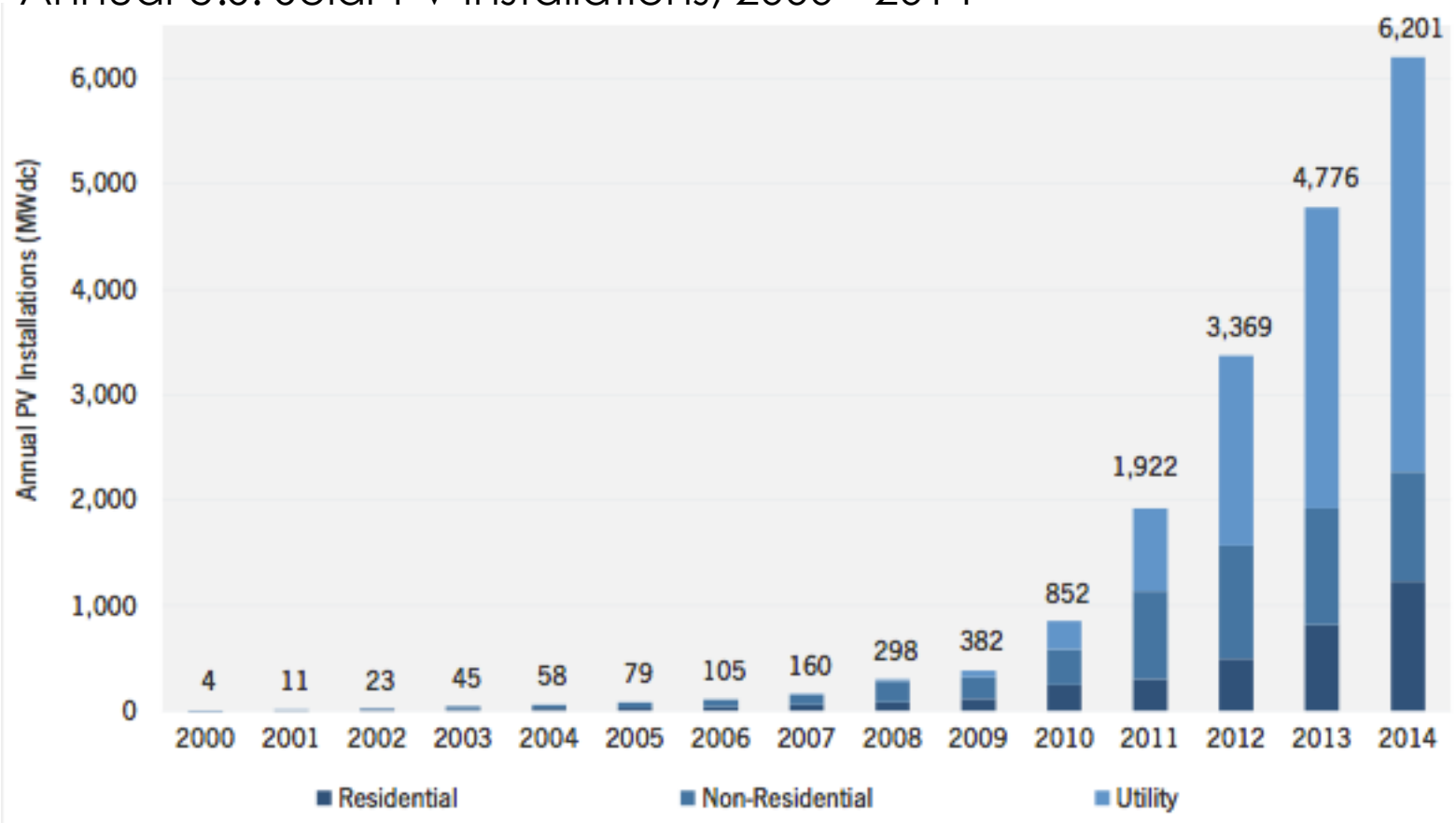
Harrison Fell, *Assistant Professor, Colorado School of Mines*

Ben Sigrin, *Energy Systems Analyst, NREL*

International BE4 Workshop, London, UK
April 21, 2015

U.S. Solar Market is Maturing

Annual U.S. Solar PV Installations, 2000 - 2014



Source: SEIA (2015)

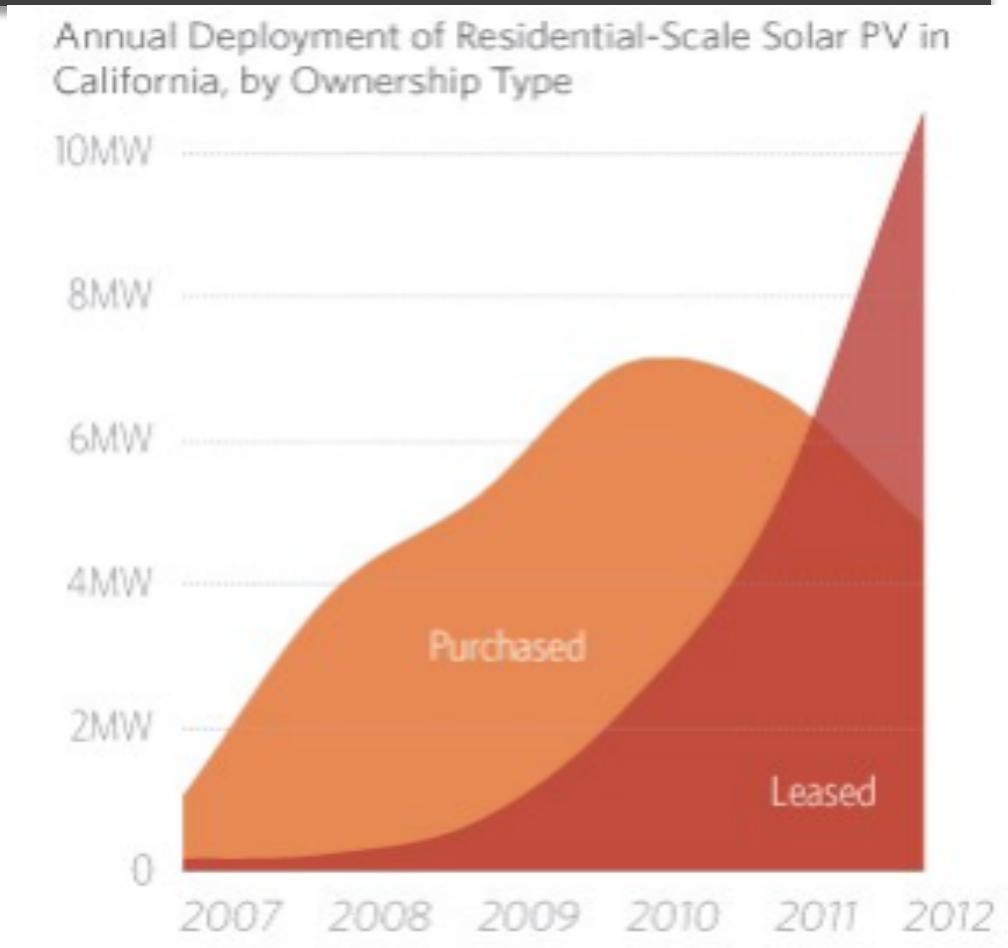
But, the solar industry is not in the clear…

"It's not like the gates of heaven open up when solar becomes cheaper."

~ Isaac Moriwake, Earthjustice

The Leasing Model

- Leasing option has helped to break down certain barriers
- It is now dominant model in the U.S. – shift from 10% of CA homeowners going solar through leasing to **over 75% in 2012**
- **One barrier that is still prominent:** customer acquisition costs remain high



New California Solar Leasing Contracts vs New California Solar Panel Purchases. Credit: Climate Policy Initiative

Source: Climate Policy Initiative (2013)

Objective

Goal

Are buyers and leasers different customer segments?

Motivation

Implications of preferences for marketing and policy

Approach

Econometric estimation using survey data

Challenge

Selection bias in the decision to buy or lease

Motivation

- Current literature on solar adoption focuses on drivers of diffusion – particularly peer effects and information networks
  - Bollinger & Gillingham (2012), Richter (2013) Graziano & Gillingham (2014)
  - Overall, peer effects are found to be significant drivers of adoption.
- Still need to better understand other behavioral drivers
- **No empirical studies exist on the decision to buy or lease**
  - Drury et al. (2012) (correlation analysis), Rai and Sigrin (2013) (engineering model) – confounding results
  - Both suffer from selection bias – data for adopters only
- **Why care?** Reducing customer acquisition costs – marketing and policy implications

Main Research Question

- **Generally:** are buyers and leasers different customer segments, exhibiting different preferences?

- **Specifically:** do buyers and leasers exhibit different information searching behavior?

- **Methodology**
  - Econometric estimation
  - Identification issues (selection bias)
  - Bivariate probit model with sample selection

Data

Combines stated and revealed preference data:

- Survey of San Diego county homeowners
- Adopters from 2007 to 2013 (1,234) and non-adopters (790) across roughly 30 zip codes
- Total of ~60 questions
  - Demographics and socioeconomic factors
  - What prompted initial interest
  - Time spent researching different components
  - Motivations for adopting – importance of various factors
- Matched to California Solar Initiative data for location information, solar system attributes, etc.

First glance: demographics and what prompted initial interest in solar PV

H0: μbuyers = μleasers *Unequal Var. Assumed*	Buyers Mean	Leasers Mean	t
Age at adoption (years)	56.8	56.3	0.647
Edu (years post-secondary)	4.64	4.23	2.91 **
Income ($1,000)	168.4	155.2	1.55
Married	0.888	0.842	1.85 (.)
Retired	0.448	0.382	1.88 (.)
Years expect to be in home	22.82	21.1	1.86 (.)
Prompted to adopt PV due to electricity rate increases	0.36	0.44	-2.22 *
Prompted to adopt PV due to upcoming remodel	0.12	0.06	3.24 ***
Prompted to adopt PV by a solar company	0.07	0.07	0.24
Prompted to adopt PV by an advertisement	0.08	0.15	-3.27 ***
Prompted to adopt PV by direct marketing	0.16	0.19	-1.20

Significance codes: *** significant at 0.1% level, ** significant at 1% level, * significant at 5% level, and (.) at 10% level

First glance: information searching

H0: μbuyers = μleasers *Unequal Var. Assumed*	Buyers Mean	Leasers Mean	t
Time researching costs	2.75	2.56	1.82 (.)
Time researching equipment	1.87	1.92	-0.55
Time researching home modifications	1.75	1.76	-0.17
Time researching fin. returns	2.24	2.09	1.55
Quotes sought for both models	0.04	0.27	-8.59 ***

Significance codes: *** significant at 0.1% level, ** significant at 1% level, * significant at 5% level, and (.) at 10% level

Identification Strategy

Objective is to regress the decision to buy (y=1) or lease (y=0) on individual-level independent variables.

But…

- **Selection bias**
  - Interdependency between technology adoption decision and business model decision
  - Leaving out 'selection' introduces bias (relevant information is omitted)
  - Standard selection models (Heckman) aren't applicable
- Unobservables that change over **time**
  - Technological advances
  - Leasing model availability
  - 'Trialability', untested commodity – uncertainty, and risk perception
- Unobservables that change across **space/location**
  - Preferences
  - Marketing campaigns

Empirical Specification:
Bivariate Probit with Sample Selection

- **Estimate two probit equations with correlated error terms**
  - Selection equation: decision to adopt
  - Outcome equation: to buy or lease

$$y_1 = x_1 \beta_1 + \varepsilon_1 \qquad \varepsilon_{1i} = \eta_i + \mu_{1i}$$
$$y_2 = x_2 \beta_2 + \varepsilon_2 \qquad \varepsilon_{2i} = \eta_i + \mu_{2i}$$

- **Independent variables:** demographics, attitudes, individual-level controls, adoption interest prompts, what is important to the adopter, etc.

- **Time (year) fixed effects and zip code level fixed effects**

- **Dependent variable in outcome equation is only partially observed**

$$y_1 = 0 \qquad Pr(y_1 = 0) = \Phi(-x_1 \beta_1)$$
$$y_1 = 1, y_2 = 0 \qquad Pr(y_1 = 1, y_2 = 0) = \Phi(x_1 \beta_1) - \Phi_2(x_1 \beta_1, x_2 \beta_2, \rho)$$
$$y_1 = 1, y_2 = 1 \qquad Pr(y_1 = 1, y_2 = 1) = \Phi_2(x_1 \beta_1, x_2 \beta_2, \rho)$$

- **Estimation is done by maximum likelihood in one step**

$$\begin{aligned} \ln L = \sum_{i=1}^{N} \{ & y_{i1} y_{i2} \ln \Phi_2(x_1 \beta_1, x_2 \beta_2; \rho) \\ & + y_{i1}(1 - y_{i2}) \ln[\Phi(x_1 \beta_1) - \Phi_2(x_1 \beta_1, x_2 \beta_2; \rho)] \\ & + (1 - y_{i1}) \ln \Phi(-x_1 \beta_1) \} \end{aligned}$$

Results: Bivariate Probit with Sample Selection Model

Variable	Coefficient	Std. error	Coefficient	Std. error
	Buy v Lease		***Adoption Decision***	
Time researching costs	**-0.238 ****	0.121	Income	0.140 ***	0.034
Time researching equipment	0.127	0.212	Years in home	0.009 *	0.005
Time researching home mods	**-0.424 *****	0.154	Married	0.211 *	0.110
Time researching fin. returns	**0.615 *****	0.163	Rate increase expectations	0.198 ***	0.051
Quotes sought for both models	**-0.960 ****	0.387	Education	-0.063	0.039
Imp. of home value	-0.057	0.177	Age	-0.002	0.005
Imp. of electricity costs	**0.405 ***	0.226	HH Age	0.000	0.003
Imp. of electricity price increases	-0.342	0.221	Retired	-0.084	0.124
Imp. of environment	0.034	0.152	HH size (sqft)	0.000	0.000
Imp. of being able to sell home	0.177	0.127	AC	0.335 ***	0.115
Monthly savings as decision metric	**-0.912 *****	0.278	Pool	0.535 ***	0.106
Rate increase expectations	0.081	0.577	Political views	-0.024	0.033
Perceived as highest savings option	**0.482 *****	0.107	Month savings as dec metric	-0.012	0.130
Married	0.709	0.669			
Education	0.199	0.220			
AC	0.887	1.048			

Number of observations: 879 Number of censored observations: 512

Rho = 0.034 Wald test (rho = 0), Prob > chi2 = 0.994

Zip code FEs, Year FEs, Errors clustered on zip code Log pseudolikelihood: -630.0661

* p<.1; ** p<.05; *** p<.01

Other variables included in buy v lease regression but not significant:

- What prompted adoption: remodeling, elect rate increases, solar company, advertising, marketing
- Retired, age, income, pool, age of house, size of house (sqft)

Without selection bias correction (univariate probit)

Variable	Coefficient	Std. error	Difference from selection model
Time researching costs	-0.133	0.102	**Significant in selection model**
Time researching equipment	-0.054	0.185	**Changed signs**
Time researching home mods	-0.366 ***	0.012	~same
Time researching fin. returns	0.496 ***	0.133	~same
Quotes sought for both models	-1.047 ***	0.352	**Less significant in selection model**
Imp. of home value	-0.004	0.160	~same
Imp. of electricity costs	0.298	0.200	**Significant in selection model**
Imp. of electricity price increases	-0.308	0.205	~same
Imp. of environment	0.039	0.099	~same
Imp. of being able to sell home	0.127	0.120	~same
Monthly savings as decision metric	-0.701 ***	0.252	~same
Elect. rate increase expectations	-0.003	0.094	~same
Perceived as highest savings option	0.522 ***	0.093	~same
Married	0.712 **	0.355	**Not significant in selection model**
Education	0.224 *	0.135	**Not significant in selection model**
AC	0.761 **	0.344	**Not significant in selection model**
No. of observations	344		
Log pseudolikelihood	-124.63512		
Time (annual) FEs, Zip code FEs, Errors clustered on zip code			
Significance codes: * p<.1; ** p<.05; *** p<.01			

Other variables included in the buy v lease regression but not significant:

- What prompted adoption: remodeling, elect rate increases, solar company, advertising, marketing
- Size of house, AC, pool, income, education, age, retired, years in home, age of house

Conclusions & Main Contributions

- **Correcting for selection bias**: Application of appropriate method for modeling the non-random selection mechanism
- Buyers and leasers exhibit **different information searching behavior**
- Insights for marketers designing strategies to increase referrals and reduce customer acquisition costs
- **Increasing data availability today allows us to better understand how decisions actually are being made – rational v. realistic agents**
- How do we integrate insights like these into optimization models?

Thank you! Questions and feedback?

jacquelyn.pless@nrel.gov